Queensland University of Technology
Brisbane Australia

Discrete-time Inverse Optimal Control with Partial-State Information: A Soft-Optimality Approach with Constrained State Estimation

Timothy L. Molloy, Dorian Tsai, Jason J. Ford, and Tristan Perez

Abstract— **In this paper, we consider the problem of estimating the parameters of an optimal control objective function based on measurements of the closed loop system. In contrast to previous work on inverse optimal control, we consider measurements that are noise-corrupted and contain only partial-state information. We propose an inverse optimal control method based on a new soft-optimality constrained methodology of state estimation. We establish a sufficient condition for recovery of the unknown objective function parameters given complete-state information, and develop results characterising the performance of our method for linear systems. We illustrate our proposed soft-optimality approach through simulations of a nonlinear and fully-actuated mechanical system.**

I. Introduction

The problem of inverse optimal control arises in many areas of study in science and engineering [1]–[13]. In particular, inverse optimal control has been successfully applied in the analysis of biological systems such as human locomotion [1]–[4], human-posture control [5], and human-controlled aircraft motion [9]. Despite its frequent use in experimental applications, the underlying theory of inverse optimal control in discrete-time systems with partial-state information has received limited attention (cf. [5] and [6]). In this paper, we propose a novel inverse optimal control method for discrete-time systems with noisy partial-state information.

Inverse optimal control is the problem of estimating the unknown objective function (or alternatively the objective function parameters) of an optimal control problem from measured optimal state and control trajectories [7], [12]. Early treatments of inverse optimal control focused on the linear quadratic regulator (LQR) problem—see [8, Section 10.6] and references therein. These problems continue to receive attention to date [7]. Recent inverse optimal control efforts have also focused on (discrete-time) Markov decision processes (MDPs) [12]–[14] that locally solve optimal control problems [14], [15]. These techniques for MDPs become intractable for continuous state and action spaces [15].

Outside of MDPs, Mombaur et al. [1] proposed a bilevel (or nested) optimisation approach for continuous-time inverse optimal control with partial-state measurements of general nonlinear systems. This bilevel approach involves solving the optimal control problem repeatedly for proposed candidate objective functions (during a numerical optimisation) [1]. It is therefore computationally expensive, particularly when solving the optimal control problem is nontrivial.

Recently, methods of inverse optimal control that avoid solving optimal control problems have been proposed on the basis of the Karush-Kuhn-Tucker (KKT) conditions in discrete-time (cf. [3], [10], [16]), and Pontryagin's minimum principle and the Hamilton-Jacobi-Bellman equation in continuous-time (cf. [2], [5], [6]). In particular, [10] exploited KKT conditions to propose an inverse optimal control method in discrete-time with observed full-state trajectories corrupted by unknown but bounded noise. However, inverse optimal control with (unbounded) noisy partial-state measurements remains an open problem [5], [6]. Furthermore, there appears to be few theoretical results establishing conditions on the system dynamics that ensure that the parameters of the objective function are identifiable (even with noise-free state information) [5]. A notable exception is [4], where identifiability conditions are established for continuous-time differentially flat systems given noise-free state measurements.

The main contribution of this paper is the proposal of a novel method of discrete-time inverse optimal control that explicitly handles noisy partial-state information—although we assume that the unknown optimal controller has access to complete-state information. Our proposed approach is based on a new method of state estimation that softly constrains state estimates to be solutions to an optimal control problem. Although our approach is inspired by the bilevel approach of [1], we avoid repeatedly solving candidate optimal control problems by exploiting optimality conditions analogous to the discrete-time minimum principle (but derived from KKT conditions). A secondary contribution of this paper is the development of a sufficient condition for the unknown parameters of the objective function to be identifiable given perfect state information. Our identifiability results hold for general discrete-time nonlinear systems, and are similar to the continuous-time results of [4] and [5].

The rest of this paper is structured as follows. In Section II, we pose our inverse optimal control problem. In Section III, we propose a soft-optimality constrained approach to inverse optimal control. In Section IV, we establish an identifiability result and partial state results for linear systems. In Section V, we present a simulation case study of a nonlinear system. We provide conclusions in Section VI.

II. Problem Formulation

Consider the deterministic discrete-time system

$$x_{k+1} = f(x_k, u_k), \quad x_0 = \bar{x} \tag{1}$$

The authors are with the School of Electrical Engineering and Computer Science, Queensland University of Technology (QUT), Brisbane, QLD 4000, Australia {t.molloy, dy.tsai, j2.ford, tristan.perez}@qut.edu.au This work was supported by an Australian Research Council Discovery Grant (DP140100896).

for $0 \leq k \leq T-1$ where $f(\cdot,\cdot) : \mathbb{R}^n \times \mathbb{R}^m \mapsto \mathbb{R}^n$ is a given (possibly nonlinear) differentiable time-invariant function, $x_k \in \mathbb{R}^n$ are the state variables and $u_k \in \mathbb{R}^m$ are the control inputs. Let us define the objective function

$$V_T\left(x_{0T}, u_{0T-1}, \alpha\right) \triangleq \sum_{k=0}^{T-1} L\left(x_k, u_k, \alpha\right),$$

with parameter $\alpha \in S_\alpha \subset \mathbb{R}^N$ where $x_{0T} \triangleq \{x_0, x_1, \ldots, x_T\}$ and $u_{0T-1} \triangleq \{u_0, u_1, \ldots, u_{T-1}\}$ are the state and control trajectories, respectively. We assume that the stage objective function $L(\cdot,\cdot,\cdot)$ is the linear combination of N (known) differentiable scalar functions $L_i(\cdot,\cdot)$, namely,

$$L\left(x_k, u_k, \alpha\right) \triangleq \sum_{i=1}^{N} \alpha_i L_i\left(x_k, u_k\right).$$

In the finite horizon optimal control problem, we are given the parameters $\alpha^* \in S_\alpha$ and the initial state $\bar{x} \in \mathbb{R}^n$, and we solve the optimisation problem:

$$\begin{aligned} \inf_{u_{0T-1}} \quad & V_T\left(x_{0T}, u_{0T-1}, \alpha^*\right) \\ \text{s.t.} \quad & x_{k+1} = f(x_k, u_k),\ k = 0, 1, \ldots, T-1 \\ & x_0 = \bar{x}. \end{aligned} \tag{2}$$

In this paper, we shall assume the existence of a (possibly nonunique) control trajectory u^*_{0T-1} solving (2) for α^* and $\bar{x}$. We use x^*_{0T} to denote an optimal state trajectory with initial state $\bar{x}$ and associated control trajectory u^*_{0T-1}.

In the inverse optimal control problem, we are interested in recovering the true (unknown) parameters $\alpha^* \in S_\alpha$ given optimal state x^*_{0T} and control u^*_{0T-1} trajectories, knowledge of the system dynamics $f(\cdot,\cdot)$, and knowledge of the functions $L_i(\cdot,\cdot)$ for $1 \leq i \leq N$. We shall further assume that any observed optimal state trajectories x^*_{0T} are only partially observed through the noise-corrupted measurements

$$y_k \triangleq h\left(x^*_k\right) + w_k, \tag{3}$$

for $0 \leq k \leq T$ where $w_k \in \mathbb{R}^l$ is a sequence of (possibly non-Gaussian) zero-mean independently and identically distributed (i.i.d.) random variables, and $h(\cdot) : \mathbb{R}^n \mapsto \mathbb{R}^l$ is a known function. Using knowledge of the system dynamics $f(\cdot,\cdot)$ and the functions $L_i(\cdot,\cdot)$ for $1 \leq i \leq N$, our aim is then to recover the unknown parameters $\alpha^* \in S_\alpha$ based on the measurements y_{0T} and controls u^*_{0T-1}.

Although we have access to the measurements y_{0T} from (3) for inverse optimal control, we assume that the unknown optimal controller has access to the complete (noise-free) states. In practical settings, such as in biological experiments, this assumption appears to be a reasonable conjecture since the (unknown) controller often has a different state measurement process than the process generating measurements for inverse optimal control.

III. Inverse Optimal Control Approaches

In this section, we present our proposed method for estimating α^* from the measurements y_{0T} and controls u^*_{0T-1} given knowledge of the dynamics $f(\cdot,\cdot)$ and functions $L_i(\cdot,\cdot)$. We first describe a bilevel (or nested) optimisation formulation of inverse optimal control with partial-state information.

A. Bilevel Optimisation Approach

Let us define the measurement cost function as

$$J_T\left(x_{0T}\right) \triangleq \sum_{k=0}^{T} \left\| y_k - h\left(x_k\right) \right\|^2 \tag{4}$$

where y_k and $h(x_k)$ are the observed and predicted measurements respectively. Here, $\|\cdot\|$ denotes the ℓ^2-vector norm. Similar to the method described in [1] for the continuous-time case, we consider a bilevel optimisation approach to inverse optimal control. This approach involves estimating the unknown trajectory x^*_{0T} and parameters $\alpha^* \in S_\alpha$ by solving the optimisation problem

$$\inf_{\alpha \in S_\alpha} \inf_{x_{0T} \in \mathcal{X}^*(\alpha)} \quad J_T\left(x_{0T}\right), \tag{5}$$

where we define $\mathcal{X}^*(\alpha)$ as the set of all optimal state sequences x_{0T} obtained by solving (2) with $\alpha^* = \alpha \in S_\alpha$. The inner subproblem of optimising over x_{0T} in (5) is therefore a constrained state estimation problem where the constraint $x_{0T} \in \mathcal{X}^*(\alpha)$ is defined with reference to the optimal control problem (2). We highlight that simultaneously estimating the states and parameters is important here since we have incomplete-state information, and estimating the states without optimality constraints may lead to a trajectory that optimises (4), but does not solve the optimal control problem (2).

Unfortunately, solving the optimisation problem (5) is nontrivial. As in [1], we could solve (5) by repeatedly solving the optimal control problem (2) for candidate parameters $\alpha \in S_\alpha$ in order to find optimal state sequences $x_{0T} \in \mathcal{X}^*(\alpha)$. We propose, however, an alternative and more tractable approach by relaxing the optimality constraint $x_{0T} \in \mathcal{X}^*(\alpha)$ in (5) using necessary conditions for optimality under (2).

B. Necessary Conditions For Optimality

In order to present the discrete-time necessary conditions for optimality derived in [17, Section 3.3], let us define the Hamiltonian of (2) for any $\alpha \in S_\alpha$ as the function

$$H\left(x_k, u_k, \lambda_k, \alpha\right) \triangleq L\left(x_k, u_k, \alpha\right) + \lambda_k' f\left(x_k, u_k\right) \tag{6}$$

for $0 \leq k \leq T-1$ where $\lambda_k \in \mathbb{R}^n$ are adjoint (or costate) variables. Let us also define the column vectors of partial derivatives of $H(x_k, u_k, \lambda_k, \alpha)$ with respect to x_k and u_k (and evaluated at x_k and u_k) as $\nabla_x H(x_k, u_k, \lambda_k, \alpha) \in \mathbb{R}^n$ and $\nabla_u H(x_k, u_k, \lambda_k, \alpha) \in \mathbb{R}^m$, respectively. We then have

$$\begin{aligned} &\nabla_x H\left(x_k, u_k, \lambda_k, \alpha\right) \\ &\quad = \sum_{i=1}^{N} \alpha_i \nabla_x L_i\left(x_k, u_k\right) + \nabla_x f\left(x_k, u_k\right) \lambda_k \end{aligned} \tag{7}$$

with

$$\nabla_x f(x_k, u_k) = \begin{bmatrix} \frac{\partial f^1(x_k,u_k)}{\partial x_k^1} & \cdots & \frac{\partial f^n(x_k,u_k)}{\partial x_k^1} \\ \vdots & \ddots & \vdots \\ \frac{\partial f^1(x_k,u_k)}{\partial x_k^n} & \cdots & \frac{\partial f^n(x_k,u_k)}{\partial x_k^n} \end{bmatrix}.$$

We note that $\nabla_u H(x_k, u_k, \lambda_k, \alpha)$ can similarly be expressed using α, $f(x_k, u_k)$, and the functions $L_i(x_k, u_k)$.

From [17, Section 3.3], if x_{0T}^* and u_{0T-1}^* are solutions to the optimal control problem (2) for a given α^*, then: (i) the trajectories x_{0T}^* and u_{0T-1}^* satisfy (1) for all $0 \leq k \leq T-1$; (ii) there exists adjoint variables $\lambda_k^* \in \mathbb{R}^n$ for $0 \leq k \leq T-1$ satisfying the backwards recursion

$$\lambda_{k-1}^* = \nabla_x H(x_k^*, u_k^*, \lambda_k^*, \alpha^*)$$

for $1 \leq k \leq T-1$ with the boundary condition $\lambda_{T-1}^* = \underline{0}$, and; (iii) the controls u_k^* are stationary points of the Hamiltonian in the sense that

$$\nabla_u H(x_k^*, u_k^*, \lambda_k^*, \alpha^*) = \underline{0} \tag{8}$$

for all $0 \leq k \leq T-1$. Here, $\underline{0}$ is a matrix of zeros with appropriate dimensions. We are now in a position to propose a soft-optimality inverse optimal control approach.

C. Proposed Soft-Optimality Approach

We propose using the condition on the gradient of the Hamiltonian (8) to replace the optimality constraint in (5) with a soft-optimality constraint. Let us define the penalty function

$$G_T\left(x_{0T-1}, \alpha\right) \triangleq \sum_{k=0}^{T-1} \left\|\nabla_u H(x_k, u_k^*, \lambda_k, \alpha)\right\|^2.$$

Here, we hide the dependence of $G_T\left(x_{0T-1}, \alpha\right)$ on the adjoint variables λ_k since we shall later express them as functions of the state sequence x_{0T-1} and parameters α. We propose combining the penalty function $G_T\left(x_{0T-1}, \alpha\right)$ with the measurement cost function (4) to form the soft-optimality cost function

$$\mathcal{J}_T(x_{0T}, \alpha) \triangleq J_T(x_{0T}) + G_T\left(x_{0T-1}, \alpha\right).$$

Now, given the state measurements y_{0T} and the optimal control trajectory u_{0T-1}^*, our proposed soft-optimality method of inverse optimal control is to solve the optimisation problem

$$\begin{aligned} \inf_{x_{0T}, \alpha} \quad & \mathcal{J}_T(x_{0T}, \alpha) \\ \text{s.t.} \quad & x_{k+1} = f(x_k, u_k^*), \ k = 0, \ldots, T-1 \\ & \lambda_{k-1} = \nabla_x H(x_k, u_k^*, \lambda_k, \alpha), \ k = 1, \ldots, T-1 \\ & \lambda_{T-1} = \underline{0} \\ & \alpha \in S_\alpha. \end{aligned} \tag{9}$$

We highlight that due to the equality constraints in (9), our method of soft-optimality reduces to finding an estimate of the initial state x_0^* and an estimate of the parameters α^*.

Before we examine the properties of our proposed soft-optimality method (9), we highlight its simplicity. Firstly, in contrast to the bilevel formulation of inverse optimal control (5), our soft-optimality approach (9) avoids solving optimal control problems (2) (which significantly reduces the complexity of solving the inverse optimal control problem). Furthermore, the subproblem of optimising over α in (9) (given a state sequence x_{0T}) is convex when S_α is a convex set. To see that optimising over α in (9) given a state sequence x_{0T} is a convex problem, we first note that the adjoint constraint equations in (9) can be rewritten as linear functions of α using (7) in the sense that

$$\begin{aligned} \lambda_{k-1} &= \sum_{i=1}^{N} \alpha_i \nabla_x L_i(x_k, u_k^*) + \nabla_x f(x_k, u_k^*) \lambda_k \\ &= \sum_{i=1}^{N} \alpha_i \lambda_{k-1}^i \end{aligned} \tag{10}$$

for $1 \leq k \leq T-1$ where the second line follows by exploiting the (backwards) recursive relationship between adjoint variables with $\lambda_{T-1} = \underline{0}$ and defining

$$\lambda_k^i \triangleq \sum_{\ell=k+1}^{T-1} \prod_{j=k+1}^{\ell-1} \nabla_x f\left(x_j, u_j^*\right) \nabla_x L_i(x_\ell, u_\ell^*)$$

for all $0 \leq k \leq T-2$ and all $1 \leq i \leq N$ with the convention $\prod_{j=k+1}^{k} \nabla_x f(x_j, u_j) \triangleq I$—where I denotes the identify matrix. We then note that the objective function $\mathcal{J}_T(x_{0T}, \alpha)$ is a convex function of α since $G_T\left(x_{0T-1}, \alpha\right)$ is a convex function of α as a result of $\nabla_u H(x_k, u_k^*, \lambda_k, \alpha)$ being a linear function of α in the sense that

$$\nabla_u H(x_k, u_k^*, \lambda_k, \alpha) = \sum_{i=1}^{N} \alpha_i \nabla_u H_k^i \tag{11}$$

for all $0 \leq k \leq T-1$ where we define the shorthand

$$\nabla_u H_k^i \triangleq \nabla_u L_i(x_k, u_k^*) + \nabla_u f(x_k, u_k^*) \lambda_k^i$$

for all $1 \leq i \leq N$.

We next examine further properties of our proposed soft-optimality approach (9).

IV. Identifiability and Linear System Results

In this section, we establish identifiability and other properties of our proposed soft-optimality method of inverse optimal control (9).

A. Identifiability of α^* and the Full State Information Case

In general, the true (unknown) parameters α^* will be a nonunique solution to our soft-optimality problem (9). For example, by inspecting (2), we note that $\alpha = r\alpha^*$ for all $r \geq 0$ will solve (9). Furthermore, the problem may be ill-posed for short time horizons T, degenerate system dynamics $f(\cdot, \cdot)$, and poor initial conditions $\bar{x}$ (such as equilibrium points that would lead to uninformative trajectories and measurements). These issues can occur even with perfect state information. The following concept of identifiability is useful for describing cases where α^* is recoverable (up to an unknown scaling factor $0 < r < \infty$) by solving (9) given complete-state information.

Definition 4.1 (Identifiability): For a given set S_α, we shall say that α^* is identifiable under the soft-optimality cost function $\mathcal{J}_T(x^*_{0T}, \alpha)$ when

$$\mathcal{J}_T(x^*_{0T}, \alpha) = 0$$

if and only if $\alpha = r\alpha^*$ for some $0 < r < \infty$ where λ_{0T-1} are given by (10) with states x^*_{0T} and parameters α.

We now establish a sufficient condition for α^* to be identifiable given complete state information.

Theorem 1: Suppose that $\alpha^* \in \mathbb{R}^N$ and that there exists some $0 < r < \infty$ such that $r\alpha^* \in S_\alpha$ where $S_\alpha = \left\{\alpha \in \mathbb{R}^N : \alpha_1 = 1\right\}$. Furthermore, suppose that the measurements provide full state information in the sense that $y_k = x^*_k$ for all $0 \le k \le T$, and define

$$\xi_T \triangleq \begin{bmatrix} 1 & \dots & 0 \\ \nabla_u H^{1*}_0 & \dots & \nabla_u H^{N*}_0 \\ \vdots & \ddots & \vdots \\ \nabla_u H^{1*}_{T-1} & \dots & \nabla_u H^{N*}_{T-1} \end{bmatrix} \in \mathbb{R}^{(mT+1)\times N}$$

where $\nabla_u H^{i*}_k$ denotes $\nabla_u H^i_k$ evaluated with $x_k = x^*_k$ for $1 \le i \le N$ and $0 \le k \le T-1$. Then solving (9) for α is equivalent to solving the system of linear equations

$$\xi_T \alpha = \left[1, 0, \dots, 0\right]',$$

and α^* is identifiable in the sense of Definition 4.1 if $\operatorname{rank}(\xi_T) = N$.

Proof: Since $y_k = x^*_k$ for all $0 \le k \le T$ we have that $J_T(x^*_{0T}) = 0$. Hence, $\mathcal{J}_T(x^*_{0T}, \alpha) = G_T\left(x^*_{0T-1}, \alpha\right)$ for all $\alpha \in S_\alpha$. From (11), we note that finding an $\alpha \in S_\alpha$ such that

$$G_T\left(x^*_{0T-1}, \alpha\right) = 0$$

under the constraints in (9) is equivalent to the problem of solving the system of linear equations $\xi_T \alpha = \left[1, 0, \dots, 0\right]'$. The proof is completed by noting that the system of equations will have a unique solution $\alpha = r\alpha^* \in S_\alpha$ when $\operatorname{rank}(\xi_T) = N$. ∎

We may interpret the rank condition of Theorem 1 as a persistence of excitation condition since it ensures that the state and control trajectories provide enough information about the parameters α^*. Although this rank condition is non-constructive, it is testable since ξ_T is independent of α^*. Finally, we note that Theorem 1 applies generally to both linear and nonlinear systems since we impose minimal conditions on the underlying system dynamics. We next develop results that only apply to linear systems.

Remark 1: In Theorem 1, the choice of the set S_α is nonunique. For example, S_α could constrain $\alpha_i = 1$ for any $1 \le i \le N$ and a rank condition similar to that in Theorem 1 would hold.

B. Specialisation to Linear Systems

We now establish results for the linear dynamics

$$f(x_k, u_k) = Ax_k + Bu_k \tag{12}$$

for $0 \le k \le T-1$ and the linear measurement model

$$h(x_k) = Cx_k \tag{13}$$

for all $0 \le k \le T$. Here, $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, and $C \in \mathbb{R}^{l\times n}$ are (known) real valued matrices. For $k \ge 1$, let us also define the controllability matrix $\mathcal{C}_k \triangleq \left[B, AB, \dots, A^{k-1}B\right] \in \mathbb{R}^{n\times mk}$, and let

$$\mathcal{O}_k \triangleq \begin{bmatrix} C \\ \vdots \\ CA^k \end{bmatrix}, \quad \mathcal{U}^*_k \triangleq \begin{bmatrix} u^*_{k-1} \\ \vdots \\ u^*_0 \end{bmatrix}, \quad \text{and} \quad \mathcal{Y}_k \triangleq \begin{bmatrix} y_0 \\ \vdots \\ y_k \end{bmatrix}$$

be the observability matrix, vector of controls, and vector of measurements, respectively. Finally, let us define the matrix

$$\mathcal{N}_T \triangleq \begin{bmatrix} \underline{0} \\ C\mathcal{C}_1\mathcal{U}^*_1 \\ \vdots \\ C\mathcal{C}_T\mathcal{U}^*_T \end{bmatrix} \in \mathbb{R}^{l(T+1)}.$$

Our first linear system result establishes that our soft-optimality method (9) can achieve accurate inverse optimal control on the basis of partial state information when an observability condition holds.

Theorem 2: Suppose that $\alpha^* \in \mathbb{R}^N$ and that there exists some $0 < r < \infty$ such that $r\alpha^* \in S_\alpha$ where $S_\alpha = \left\{\alpha \in \mathbb{R}^N : \alpha_1 = 1\right\}$. Furthermore, suppose that $f(\cdot, \cdot)$ is given by (12), and that the measurements y_k contain noise-free partial state information in the sense that $y_k = h(x^*_k)$ for all $0 \le k \le T$ where $h(\cdot)$ is given by (13). If $\operatorname{rank}(\mathcal{O}_T) = n$ and $\operatorname{rank}(\xi_T) = N$, then $(x^*_{0T}, r\alpha^*)$ is the unique solution to our soft-optimality inverse optimal control problem (9).

Proof: We note that x^*_{0T} and $r\alpha^*$ are solutions to (9) with $\mathcal{J}_T(x^*_{0T}, \alpha^*) = 0$ when $y_k = Cx^*_k$ for all $0 \le k \le T$.

To show that x^*_{0T} is the unique (globally) optimal state sequence solving (9), it suffices to show that $J_T(x_{0T}) = 0$ if and only if $x_{0T} = x^*_{0T}$ (since $G_T(\cdot, \cdot)$ in (9) is always nonnegative). From (12), we have that $x_k = A^k x_0 + \mathcal{C}_k \mathcal{U}^*_k$ for all $1 \le k \le T$ and so the condition $J_T(x_{0T}) = 0$ may be rewritten as the system of linear equations $\mathcal{Y}_T = \mathcal{O}_T x_0 + \mathcal{N}_T$. Importantly, this system involves n unknowns (i.e., the n components of the initial state $x_0 \in \mathbb{R}^n$), and so $\operatorname{rank}(\mathcal{O}_T) = n$ ensures that $x_0 = x^*_0$ is the only solution. Hence, $\operatorname{rank}(\mathcal{O}_T) = n$ implies that $J_T(x_{0T}) = 0$ if and only if $x_{0T} = x^*_{0T}$.

It follows that $r\alpha^*$ is the unique global optimiser of (9) when $\operatorname{rank}(\mathcal{O}_T) = n$ and $\operatorname{rank}(\xi_T) = N$ since Theorem 1 gives that $\mathcal{J}_T(x^*_{0T}, \alpha) = G_T\left(x^*_{0T-1}, \alpha\right) = 0$ if and only if $\alpha = r\alpha^*$. The theorem assertion then follows. ∎

The proof of Theorem 2 suggests that when the state measurements y_k are free of noise, we may sequentially estimate the state, then perform inverse optimal control. In this case, there appears to be no clear advantage in simultaneously performing state estimation and estimation of α^*. We shall later illustrate in simulations that sequentially performing state estimation and inverse optimal control when the measurements are corrupted by noise can lead to worse performance than our soft-optimality approach (9).

We now describe the link between the state and parameter estimates produced by our soft-optimality approach (9) for linear systems when the measurements y_k provide noisy partial-state information and the stage objective function is quadratic in the sense that

$$L\left(x_k, u_k, \alpha\right) = x_k' \alpha^x x_k + u_k' \alpha^u u_k. \tag{14}$$

Here, $\alpha^x \triangleq \operatorname{diag}\left(\alpha_1, \alpha_2, \ldots, \alpha_n\right) \in \mathbb{R}^{n \times n}$ is the (positive semi-definite) diagonal matrix with main diagonal $\alpha_1, \alpha_2, \ldots, \alpha_n$, and $\alpha^u \triangleq \operatorname{diag}\left(\alpha_{n+1}, \alpha_{n+2}, \ldots, \alpha_N\right) \in \mathbb{R}^{m \times m}$ is defined similarly as a (positive definite) diagonal matrix. To present our result, let us define the matrices

$$\mathcal{A}_k\left(\alpha\right) \triangleq -2B' \sum_{\ell=k+1}^{T-1} \left(A^{\ell-k-1}\right)' \alpha^x A^\ell \in \mathbb{R}^{m \times n}$$

for all $0 \leq k \leq T-2$ with $\mathcal{A}_{T-1} \triangleq \underline{0}$, and the vectors

$$\mathcal{P}_k\left(\alpha\right) \triangleq 2\alpha^u u_k^* + 2B' \sum_{\ell=k+1}^{T-1} \left(A^{\ell-k-1}\right)' \alpha^x \mathcal{C}_\ell \mathcal{U}_\ell^* \in \mathbb{R}^m$$

for all $0 \leq k \leq T-2$ with $\mathcal{P}_{T-1}\left(\alpha\right) \triangleq 2\alpha^u u_{T-1}^*$. Let us also define the stacked matrices

$$\mathcal{M}_T\left(\alpha\right) \triangleq \begin{bmatrix} \mathcal{A}_0\left(\alpha\right) \\ \vdots \\ \mathcal{A}_{T-1}\left(\alpha\right) \end{bmatrix}, \quad \mathcal{Z}_T\left(\alpha\right) \triangleq \begin{bmatrix} \mathcal{P}_0\left(\alpha\right) \\ \vdots \\ \mathcal{P}_{T-1}\left(\alpha\right) \end{bmatrix}$$

and $\tilde{\mathcal{Y}}_T \triangleq \mathcal{Y}_T - \mathcal{N}_T$. We now have the following LQR result.

Theorem 3: Suppose that $f\left(\cdot,\cdot\right)$ is given by (12), $y_k = h\left(x_k^*\right) + w_k$ for all $0 \leq k \leq T$ where $h\left(\cdot\right)$ is given by (13) and where $w_k \in \mathbb{R}^l$ for $0 \leq k \leq T$ are zero-mean i.i.d. random variables. Consider any $\alpha \in S_\alpha$ where $S_\alpha \subset \mathbb{R}^N$ is the set of all vectors such that α^x is positive semi-definite and α^u is positive definite. If the matrix $\mathcal{R}_T\left(\alpha\right) \triangleq \mathcal{O}_T' \mathcal{O}_T + \mathcal{M}_T'\left(\alpha\right) \mathcal{M}_T\left(\alpha\right) \in \mathbb{R}^{n \times n}$ is invertible, then the optimal state estimate $\hat{x}_0$ under our soft-optimality approach (9) is

$$\hat{x}_0 = \mathcal{R}_T^{-1}\left(\alpha\right) \left[\mathcal{O}_T' \tilde{\mathcal{Y}}_T + \mathcal{M}_T'\left(\alpha\right) \mathcal{Z}_T\left(\alpha\right)\right] \tag{15}$$

and satisfies

$$E\left[\hat{x}_0\right] = \mathcal{R}_T^{-1}\left(\alpha\right) \left[\mathcal{O}_T' \mathcal{O}_T x_0^* + \mathcal{M}_T'\left(\alpha\right) \mathcal{Z}_T\left(\alpha\right)\right]. \tag{16}$$

Furthermore, $E\left[\hat{x}_0\right] = x_0^*$ when $\alpha = \alpha^*$ in (16).

Proof: Consider any $\alpha \in S_\alpha$. From (12) and (14) we have that $\nabla_x f\left(x_k, u_k\right) = A'$, $\nabla_u f\left(x_k, u_k\right) = B'$, $\nabla_x L\left(x_k, u_k, \alpha\right) = 2\alpha^x x_k$, and $\nabla_u L\left(x_k, u_k, \alpha\right) = 2\alpha^u u_k$. Differentiating (6) therefore gives

$$\begin{aligned} \nabla_u H\left(x_k, u_k^*, \lambda_k, \alpha\right) &= 2\alpha^u u_k^* \\ &\quad + 2B' \sum_{\ell=k+1}^{T-1} \left(A^{\ell-k-1}\right)' \alpha^x x_\ell \\ &= \mathcal{P}_k\left(\alpha\right) - \mathcal{A}_k\left(\alpha\right) x_0 \end{aligned}$$

for all $0 \leq k \leq T-1$ since $x_\ell = A^\ell x_0 + \mathcal{C}_\ell \mathcal{U}_\ell^*$ from (12). Hence,

$$\begin{aligned} &G_T\left(x_{0T-1}, \alpha\right) \\ &\quad = \left[\mathcal{Z}_T\left(\alpha\right) - \mathcal{M}_T\left(\alpha\right) x_0\right]' \left[\mathcal{Z}_T\left(\alpha\right) - \mathcal{M}_T\left(\alpha\right) x_0\right], \end{aligned}$$

and

$$J_T\left(x_{0T}\right) = \left[\tilde{\mathcal{Y}}_T - \mathcal{O}_T x_0\right]' \left[\tilde{\mathcal{Y}}_T - \mathcal{O}_T x_0\right].$$

Differentiating $\mathcal{J}_T\left(x_{0T}, \alpha\right)$ therefore gives that

$$\begin{aligned} \nabla_{x_0} \mathcal{J}_T\left(x_{0T}, \alpha\right) &= -2\mathcal{O}_T' \left[\tilde{\mathcal{Y}}_T - \mathcal{O}_T x_0\right] \\ &\quad - 2\mathcal{M}_T'\left(\alpha\right) \left[\mathcal{Z}_T\left(\alpha\right) - \mathcal{M}_T\left(\alpha\right) x_0\right]. \end{aligned}$$

Now, (15) follows by setting $\nabla_{x_0} \mathcal{J}_T\left(x_{0T}, \alpha\right) = \underline{0}$ and rearranging (noting that $\mathcal{R}_T\left(\alpha\right)$ is invertible under the theorem conditions). Under (12) and (13), we have that $\tilde{\mathcal{Y}}_T = \mathcal{O}_T x_0^* + [w_0', \ldots, w_T']'$, and so taking the expectation of (15) gives (16). The final theorem statement $E\left[\hat{x}_0\right] = x_0^*$ when $\alpha = \alpha^*$ follows from (16) by noting that $\mathcal{Z}_T\left(\alpha^*\right) = \mathcal{M}_T\left(\alpha^*\right) x_0^*$ and by recalling the definition of $\mathcal{R}_T\left(\alpha^*\right)$. ■

Theorem 3 establishes that the initial state estimate $\hat{x}_0$ produced by our soft-optimality approach (9) is a function of α, and is unbiased when we recover the true parameters α^* (but may be biased for other values of α). Hence, if the measurements y_k are noise-corrupted, our soft-optimality approach (9) provides different results compared to a sequential process of first performing state estimation, and then estimating α^*. We will further illustrate the properties of our approach through simulations of a nonlinear system.

V. Case Study

In this section, we present simulations of a pendulum with a torque-actuated joint as the control input. The discrete-time, nonlinear dynamics of the system are

$$x_{k+1} = \begin{bmatrix} x_{1,k} + \Delta x_{2,k} \\ x_{2,k} + \Delta \left\{ -\frac{mg\ell}{J} \sin x_{1,k} - \frac{d}{J} x_{2,k} + \frac{1}{J} u_k \right\} \end{bmatrix}$$

where the two state components, $x_{1,k}$ and $x_{2,k}$, are the pendulum's angular position (in radians) and angular velocity (in radians per second), respectively. Here, the control input u_k is the torque (in Newton meters, Nm), $\Delta = 0.1$s is the incremental time step, $g = 9.81$m/s^2 is the acceleration due to gravity, and $m = 4$kg, $\ell = 0.5$m, $J = 0.15$kgm^2 and $d = 0.4$Nms/rad are the mass, length, moment of inertia, and friction in the pendulum system, respectively.

For the optimal control problem (2), we selected the stage objective function

$$\begin{aligned} L(x_k, u_k, \alpha) = (x_k - x_g)' &\begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{bmatrix} (x_k - x_g) \\ &+ \alpha_3 (u_k - mg\ell \sin x_{1,g})^2 \end{aligned} \tag{17}$$

with $x_g = (\pi/180)[-20°, 0]'$. Under this stage objective function, solving the optimal control problem (2) for a given horizon T corresponds to regulating the pendulum system to the desired goal state x_g using the control torque input u_k.

Importantly, the term subtracted from u_k in (17) ensures that x_g is an equilibrium point of the closed loop system.

For the purpose of inverse optimal control, we implemented both the bilevel approach (BL) of (5), and our soft-optimality approach (SO) of (9). We also implemented a sequential version of our soft-optimality approach (SQO) in which we first estimate the initial state x_0^* by solving the least squares problem

$$\begin{aligned} \inf_{x_0} \quad & J_T\left(x_{0T}\right) \\ \text{s.t.} \quad & x_{k+1} = f\left(x_k, u_k^*\right), \; k = 0, \ldots, T-1 \end{aligned}$$

and then use this estimated initial state $\hat{x}_0$ to solve

$$\begin{aligned} \inf_{\alpha} \quad & G_T\left(x_{0T-1}, \alpha\right) \\ \text{s.t.} \quad & x_0 = \hat{x}_0 \\ & x_{k+1} = f\left(x_k, u_k^*\right), \; k = 0, \ldots, T-1 \\ & \lambda_{k-1} = \nabla_x H\left(x_k, u_k^*, \lambda_k, \alpha\right), \; k = 1, \ldots, T-1 \\ & \lambda_{T-1} = \underline{0} \\ & \alpha \in S_\alpha \end{aligned}$$

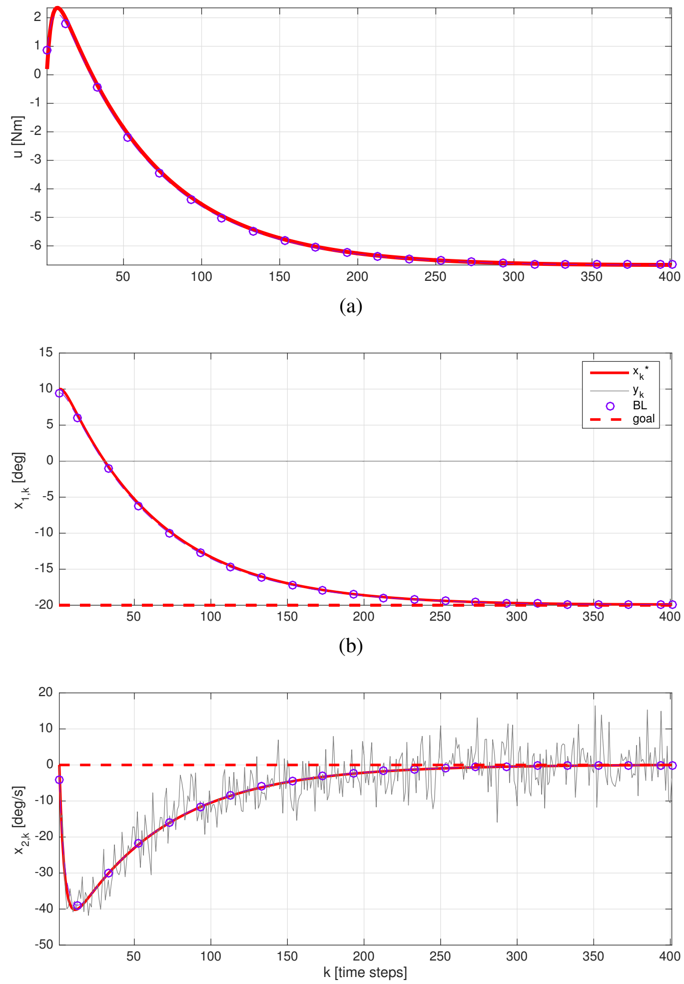
for an estimate of the unknown coefficients α^*. We solved the optimisations in the SO and SQO approaches using MATLAB's `fmincon` routine with the interior point algorithm. We selected the constraint set $S_\alpha \triangleq \left\{\alpha \in \mathbb{R}^3 : \|\alpha\| = 1 \text{ and } \alpha^i > 0 \text{ for all } 1 \leq i \leq 3\right\}$.

A. Simulation Example

To illustrate the performance of the three inverse optimal control approaches, we first simulated the true closed loop system by solving (2) with $\alpha^* = [0.9283, 0.3713, 0.0185]'$ and $T = 400$ from an initial state of $x_0^* = (\pi/180)[10°, 0]'$. We generated noisy measurements of the angular velocity using $y_k = [0, 1]x_k^* + w_k$ for $0 \leq k \leq T$ and one realisation of zero-mean Gaussian noise for w_k with variance $\sigma^2 = 0.012$ (rad/s)^2. We then applied the three inverse optimal control algorithms to estimate the initial state x_0^* and the objective function parameters α^*. To visualise the results, we compare the true optimal trajectories with those of the optimal control problem (2) solved with the estimated initial state and objective function parameters.

The optimal and measured trajectories are shown in Fig. 1 together with the reconstructed trajectories based on the estimates obtained. The estimates of the initial state $x_0^* = (\pi/180)[10°, 0]'$ and coefficients $\alpha^* = [0.9283, 0.3713, 0.0185]'$ are provided in Table I.

From Table I, we see that all methods are able to estimate the unknown the initial state and objective function parameters. For this noise realisation, our proposed soft-optimality method outperforms both the bilevel and sequential methods, and the sequential method outperforms the bilevel method. Although our method outperforms the bilevel method the results of Fig. 1 suggest that the reconstructed trajectories based on the bilevel method estimates are still close to optimal trajectories. The relatively poor performance of the bilevel method in recovering the true parameters is likely due to its cost function (5) being purely a function of the state trajectories (and only accounting for α indirectly via the constraint). In contrast, the cost functions of the sequential and soft-optimality methods are directly dependent on α.



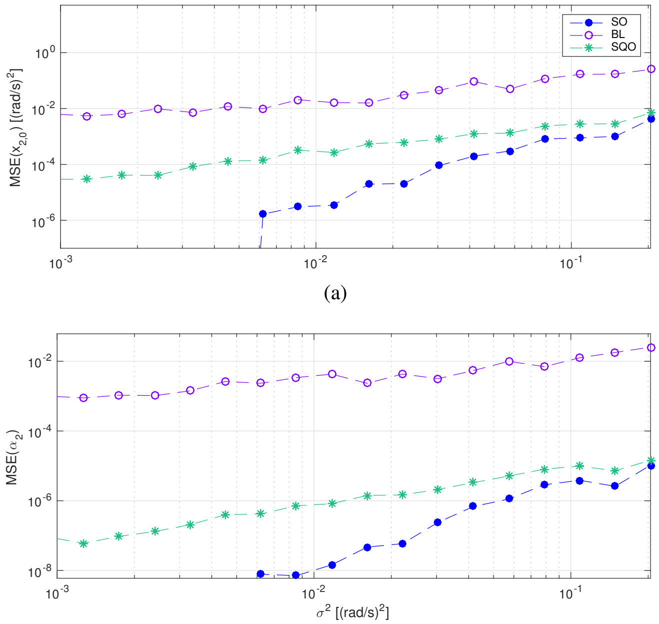
(c)

Fig. 1. Trajectories of the: (a) control input, (b) angular position, and (c) angular velocity and measurements for inverse optimal control. Starting from rest at $10°$, the pendulum is driven to $-20°$. Only the control trajectory u_{0T-1}^* and noise corrupted measurements y_k (with noise variance $\sigma^2 = 0.012$) are provided to the inverse optimal control methods. The (estimated) trajectories from the bilevel approach (BL) are superimposed. Trajectories estimated with the sequential and soft-optimality methods are not shown because they are indistinguishable from the optimal trajectories.

TABLE I

TRUE AND ESTIMATED INITIAL STATES AND PARAMETERS FOR THE BILEVEL (BL), SEQUENTIAL SOFT-OPTIMALITY (SQO), AND SOFT-OPTIMALITY (SO) APPROACHES. THE NOISE VARIANCE WAS $\sigma^2 = 0.012$ (RAD/S)^2.

Parameter	True	BL	SQO	SO
$x_{1,0}$	10.000	9.4604	9.9783	10.000
$x_{2,0}$	0.0000	-0.6238	-0.0035	0.0000
α_1	0.9283	0.9184	0.9283	0.9283
α_2	0.3713	0.3951	0.3712	0.3713
α_3	0.0185	0.0178	0.0191	0.0185



(a)

(b)

Fig. 2. Mean squared error (MSE) versus noise variance σ^2 for estimates of: (a) $x_{2,0}$, and (b) α_2 from the bilevel (BL), sequential soft-optimality (SQO), and soft-optimality (SO) approaches. The MSEs of SO for $\sigma^2 < 0.005$ (rad/s)^2 are not shown since they are zero to machine precision.

B. Effect of Noise

In order to investigate the impact of noise on the performance of the three inverse optimal control methods, we repeated our simulation example for noise variances between $\sigma^2 = 0.001$ and $\sigma^2 = 0.2$ (rad/s)^2. For each noise variance, we generated 50 noise realisations and performed inverse optimal control on the basis of the resulting noisy partial-state measurements y_{0T} and control trajectories u^*_{0T-1}. The mean squared error (MSE) in the estimates of the parameter α_2 and the initial pendulum angular velocity $x_{2,0}$ over the 50 noise realisations are shown in Fig. 2 for each of the noise variances considered. The MSEs of the estimates of $x_{1,0}$, α_1 and α_3 were better than (or similar to) those shown in Fig. 2.

The MSEs reported in Fig. 2 suggest that the performance of all three methods degrades as the noise variance increases. Indeed, the MSE of all three methods increases roughly linearly on the log-scale, which indicates a power relationship between the MSEs and the variance of the noise σ^2. Consistent with our observation in the previous simulation example, the bilevel approach appears to perform the worst in the MSE sense. Significantly, our soft-optimality method appears to offer lower MSE estimates compared to the sequential and bilevel approaches. In particular, for variances $\sigma^2 < 0.005$ (rad/s)^2, the soft-optimality MSEs are zero to machine precision (and therefore not reported in Fig. 2).

VI. Conclusion

We consider the problem of estimating the parameters of an optimal control objective function from noisy partial-state measurements of the closed loop system. We propose an inverse optimal control method based on a new soft-optimality constrained methodology of state estimation, which reduces computations and outperforms an extension (to discrete-time systems) of bilevel approaches from previous literature. We establish a sufficient condition for recovery of the unknown objective function parameters given complete-state information and we develop results characterising the performance for partial and noise-corrupted state measurements in linear systems. We illustrate our proposed soft-optimality approach through simulations of a nonlinear mechanical system. Future work will consider uncertainty in the measurement of the control trajectory.

References

[1] K. Mombaur, A. Truong, and J.-P. Laumond, "From human to humanoid locomotion—an inverse optimal control approach," *Autonomous robots*, vol. 28, no. 3, pp. 369–383, 2010.

[2] K. Hatz, J. P. Schlöder, and H. G. Bock, "Estimating parameters in optimal control problems," *SIAM Journal on Scientific Computing*, vol. 34, no. 3, pp. A1707–A1728, 2012.

[3] A.-S. Puydupin-Jamin, M. Johnson, and T. Bretl, "A convex approach to inverse optimal control and its application to modeling human locomotion," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, May 2012, pp. 531–536.

[4] N. Aghasadeghi and T. Bretl, "Inverse optimal control for differentially flat systems with application to locomotion modeling," in *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, May 2014, pp. 6018–6025.

[5] E. Pauwels, D. Henrion, and J.-B. Lasserre, "Inverse optimal control with polynomial optimization," in *Decision and Control (CDC), 2014 IEEE 53rd Annual Conference on*, Dec 2014, pp. 5581–5586.

[6] M. Johnson, N. Aghasadeghi, and T. Bretl, "Inverse optimal control for deterministic continuous-time nonlinear systems," in *Decision and Control (CDC), 2013 IEEE 52nd Annual Conference on*, Dec 2013, pp. 2906–2913.

[7] M. Priess, R. Conway, J. Choi, J. Popovich, and C. Radcliffe, "Solutions to the Inverse LQR Problem With Application to Biological Systems Analysis," *Control Systems Technology, IEEE Transactions on*, vol. 23, no. 2, pp. 770–777, March 2015.

[8] S. P. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear matrix inequalities in system and control theory*. SIAM, 1994, vol. 15.

[9] T. Maillot, U. Serres, J.-P. Gauthier, and A. Ajami, "How pilots fly: an inverse optimal control problem approach," in *Decision and Control (CDC), 2013 IEEE 52nd Annual Conference on*. IEEE, 2013, pp. 1792–1797.

[10] A. Panchea and N. Ramdani, "Towards solving inverse optimal control in a bounded-error framework," in *American Control Conference (ACC), 2015*, July 2015, pp. 4910–4915.

[11] N. Aghasadeghi, A. Long, and T. Bretl, "Inverse optimal control for a hybrid dynamical system with impacts," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, May 2012, pp. 4962–4967.

[12] P. Abbeel and A. Y. Ng, "Apprenticeship learning via inverse reinforcement learning," in *Proceedings of the twenty-first international conference on Machine learning*. ACM, 2004, p. 1.

[13] N. D. Ratliff, J. A. Bagnell, and M. A. Zinkevich, "Maximum margin planning," in *Proceedings of the 23rd international conference on Machine learning*. ACM, 2006, pp. 729–736.

[14] B. D. Ziebart, A. L. Maas, J. A. Bagnell, and A. K. Dey, "Maximum entropy inverse reinforcement learning," in *AAAI Conference on Artificial Intelligence*, 2008, pp. 1433–1438.

[15] S. Levine and V. Koltun, "Continuous inverse optimal control with locally optimal examples," in *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*, 2012, pp. 41–48.

[16] A. Keshavarz, Y. Wang, and S. Boyd, "Imputing a convex objective function," in *Intelligent Control (ISIC), 2011 IEEE International Symposium on*. IEEE, 2011, pp. 613–619.

[17] G. C. Goodwin, M. M. Seron, and J. A. De Doná, *Constrained Control and Estimation: An Optimisation Approach*. Springer Science & Business Media, 2006.